Interference of irrelevant information in multisensory selection depends on attentional set

Anne Jensen¹ · Simon Merz¹ · Charles Spence² · Christian Frings¹

Published online: 23 August 2019
© The Psychonomic Society, Inc. 2019

Abstract

In the multisensory world in which we live, certain objects and events are of more relevance than others. In the laboratory, this broadly equates to the distinction between targets and distractors. In selection situations like the flanker task, the evidence suggests that the processing of multisensory distractors is influenced by attention. Here, multisensory distractor processing was investigated by modulating attentional set in three experiments in a flanker interference task, in which the targets were unisensory while the distractors were multisensory. Attentional set was modulated by making the target modality either predictable or unpredictable (Experiments 1 vs. 2, respectively). In Experiment 3, this manipulation was implemented on a within-experiment basis. Furthermore, the third experiment compared audiovisual distractors (used in all experiments) with distractors with one feature in a neutral modality (i.e., touch), that never appeared as the target modality in the flanker task. The results demonstrate that there was no interference from the response-compatible crossmodal distractor feature when the target modality was predictable (i.e., blocked). However, when the modality was varied on a trial-by-trial basis, this crossmodal feature significantly influenced information processing. By contrast, a multisensory distractor with a neutral crossmodal feature never influenced behavior. This finding suggests that the processing of multisensory distractors depends on attentional set. When the target modality varies randomly, participants include features from both modalities in their attentional set and the irrelevant crossmodal feature, now part of the set, influences information processing. In contrast, interference from the crossmodal distractor feature does not occur when it is not part of the attentional set.

Keywords Multisensory perception · Multisensory selection · Distractor processing · Attentional set

Daily, our senses are confronted with myriad sources of information. To prevent sensory overload, mechanisms of selective attention are needed. Specifically, we have to focus on a particular stimulus or event while shielding ourselves from distraction: the relevant signals need to be (endogenously) selected against a background of other (irrelevant) signals that may interfere with our ability to achieve our current goals. In daily life, stimuli often occur in several different senses simultaneously. However, research on selection tasks with multisensory stimuli, and specifically on the processing of multisensory distractors, is scarce. Hence, the question of how multiple competing multisensory events are processed is currently poorly understood. In

two previous studies, we introduced a novel multisensory flanker task, in which the target and distractor stimuli comprised two features from different sensory modalities (Jensen, Merz, Spence, & Frings, 2019; Merz, Jensen, Spence, & Frings, 2019). In these studies, overt spatial attention was shown to modulate the multisensory integration of irrelevant stimuli. At fixation, features from different sensory modalities were integrated while they were processed at the feature-level when presented away from fixation. In the present study, we further examine attentional influences on the processing of irrelevant multisensory stimuli (i.e., distractors). Multisensory distractors are presented with unisensory stimuli while manipulating the participant's attentional set (rather than their spatial attention, as in our previous research). The attentional set (e.g., Hommel, Memelink, Zmigrod, & Colzato, 2014; Remington & Folk, 2001; Theeuwes, 1994) refers to the set of stimuli that are currently task relevant and that appear to capture attention more effectively than other sensory inputs unrelated to the participant's task. In the present study, the attentional set was modulated by varying the predictability of the target modality, and,

✉ Anne Jensen
jensen@uni-trier.de

¹ Department of Psychology, University of Trier,
D-54286 Trier, Germany

² University of Oxford, Oxford, UK

thus, the set of stimuli that can be focused on within one block of experimental trials. When the target modality is predictable, the features of the other modality do not form part of the current attentional set (i.e., the attentional set is unisensory) while the features of both modalities are included (multisensory attentional set) when it is unpredictable.

Previously, unisensory selection processes have been investigated in flanker interference tasks (e.g., Chan, Merrifield, & Spence, 2005; Driver & Grossenbacher, 1996; Eriksen & Eriksen, 1974; Gallace, Soto-Faraco, Dalton, Kreukniet, & Spence, 2008; Spence, Ranson, & Driver, 2000) and with targets and distractors that were both multisensory (audiovisual; Jensen et al., 2019; visuotactile: Merz et al., 2019). In the latter case, overt spatial attention was found to modulate multisensory distractor processing. The effects of the distractor features interacted when overt spatial attention was directed toward the distractor (i.e., when participants fixated the distractor), but were processed independently when the participants looked elsewhere. Thus, overt spatial attention modulated the integration of irrelevant multisensory information. Meanwhile, the two distractor features always elicited individual congruency effects. This result is surprising given that the target was defined by a combination of two features (and only together were they mapped onto a response). However, the two distractor features individually affected target processing, suggesting individual response overlap with the target.

Next to overt spatial attention, there are, of course, various other types of attention that might also be expected to affect multisensory distractor processing. Studies on contingent attentional capture and crossmodal congruency effects (e.g., Folk, Remington, & Johnston, 1992; Gibson & Kelsey, 1998; Hommel et al., 2014; Huang, Su, Zhen, & Qu, 2016; Ito & Kawahara, 2016; Mast & Frings, 2014; Mast, Frings, & Spence, 2014, 2015, 2017; Matusz & Eimer, 2011, 2013; Remington & Folk, 2001; Theeuwes, 1994; Wada, 2011; see also Spence, 2010) have shown that a participant's current task demands determine the processing of irrelevant (crossmodal) stimuli. Stimulus or feature dimensions that are part of the attentional set — that is, target stimuli or feature dimensions that are responded to are more likely to capture attention and interfere when they are presented as distractors in comparison to unrelated distractors. For example, a color distractor was shown to interfere with a target that was defined by color and onset, but not with an onset-only target (Gibson & Kelsey, 1998). Furthermore, Ito and Kawahara (2016) reported that the target dimension, as well as specific relevant features in that dimension, had distinct effects on attentional capture. Hence, only response relevant feature dimensions would appear to interfere. However, feature dimensions do not need to have a response assigned to them in order to be made task-relevant (Mast & Frings, 2014). For instance, Mast and Frings induced a successful manipulation of attentional

set with crossmodal stimuli by correlating a response irrelevant tactile feature with a visual target. When varied, the response irrelevant vibration was a valid cue for the identity of the target and, thus, helped participants selecting target against distractor. In this case, it interfered with the response to the visual target. When this tactile feature was kept constant, on the other hand, the crossmodal feature was ignored completely as it provided no additional information to facilitate target selection. These findings clearly demonstrate that the attentional set determines both unisensory as well as crossmodal distractor interference.

Additionally, effects of expectancy or prior knowledge have been found in other crossmodal studies (Klapetek, Ngo, & Spence, 2012; Sarmiento, Matusz, Sanabria, & Murray, 2016; Spence, Nicholls, & Driver, 2001; Wang & Theeuwes, 2018). For instance, crossmodal correspondences (such as the association between the brightness of visual stimuli and the pitch of sounds; see Spence, 2011) only produced a crossmodal congruency effect in a cueing paradigm when the presence of the cue was blocked — that is, when it was predictable, and the participants were informed about the correspondence (Klapetek, Ngo, & Spence, 2012; see also Spence, Nicholls, & Driver, 2001). Similarly, high predictability (elicited by increasing the likelihood of one specific target modality) led to speeded reactions in that modality (Spence et al., 2001). Further, distraction is modulated by the predictability of the location that the distractor appears at (e.g., Wang & Theeuwes, 2018). When a distractor was presented at one specific location with a frequency that was above chance, the interference by distractors at that location was larger than for distractors presented at a different location. Accordingly, statistical predictability is taken into account, consciously or otherwise, and influences interference effects.

Transferring this idea to the present context, modulating the predictability of the target modality, and hence, the task relevance of the nontarget modality, most likely induces modality-specific attentional sets. When blocked, the attentional set might be narrowed to just one modality while both modalities should be included when the modality of the target changes unpredictably. Critically, this induction of attentional set might affect the processing of the multisensory distractor. In the present study, the distractor contains one feature from the target modality (i.e., an intramodal feature) and one feature from the nontarget modality (i.e., a crossmodal feature). A manipulation of attentional set is thought to especially influence the effect of the crossmodal distractor feature as the task relevance of the nontarget modality is modulated. Furthermore, the crossmodal feature can, at most, elicit a response compatibility effect (distractor features interfere because they are mapped on to the same responses as the target features). As suggested by our previous studies (Jensen et al., 2019; Merz et al., 2019), the intramodal feature will probably induce a (perceptual) congruency effect based on stimulus identity.

Furthermore, several previous studies have investigated the advantage (or disadvantage) of multisensory distractors as compared to unisensory distractors in unisensory targets tasks (Fong, Hui, Fung, Chu, & Wang, 2018; Lunn, Sjoblom, Ward, Soto-Faraco, & Forster, 2019; Santangelo, Ho, & Spence, 2008; Santangelo & Spence, 2007; see Spence & Santangelo, 2009, for a review). In a cueing paradigm (e.g., Santangelo et al., 2008), different influences of multisensory and unisensory cues were revealed under a manipulation of perceptual load (see Lavie, 1995, 2005, 2010). In particular, under conditions of low perceptual load (i.e., low display complexity), both cues produced equal attentional capture effects, while under high load conditions, only multisensory cues successfully captured participants' attention. It is assumed that multisensory cues are sometimes preattentively integrated to guide the subsequent direction of spatial attention (e.g., Spence & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001). In contrast, more recent findings suggest that such effects may be dependent on task demands. When participants are asked to respond to the distractors (i.e., when they are made relevant, multisensory advantages over unisensory stimuli were found regardless of load; Lunn et al., 2019). However, load still affected multisensory processing, even if to a lesser extent than unisensory processing, thus questioning the automaticity of multisensory integration. On the other hand, when participants were asked to ignore the irrelevant input (as in the flanker paradigm), multisensory distractors did not capture attention in high load conditions and failed to produce more interference than unisensory input under low load.

Critical to the context of the present study, the distractor stimuli used in the distraction condition of Lunn and colleagues' study (2019) were not part of the attentional set (i.e., a meowing cat was used as a distractor in a letter search display). With distractors matching the attentional set, Fong et al. (2018) reported a larger congruency effect of multisensory distractors as compared with unisensory distractors in a visual flanker interference task. In the multisensory condition, an additional auditory stimulus feature was presented with the visual distractor feature. In all studies, however, the multisensory distractors provided redundant information as in, for example, the word *red* (auditory information) together with a red circle (visual information). In the present study, previously unrelated features are used as parts of the multisensory targets and distractors in order to investigate how this form of multisensory distractor processing differs from unisensory distractor processing. Moreover, the matching of the irrelevant features to the attentional set is manipulated, which is thought to affect the potential for distraction.

The present study

We report three experiments designed to investigate how multisensory distractor processing is modulated by attentional

processes in multisensory selection. In particular, we were interested in the role of attentional set in modulating flanker interference effects (Eriksen & Eriksen, 1974; see also Chan et al., 2005; Gallace et al., 2008, for auditory and tactile flanker effects) of multisensory distractors that were presented with unisensory targets. A modulation of attentional set was induced by varying the predictability of the target modality. In all three experiments, visual and auditory stimuli were used as targets. In Experiments 1 and 2, audiovisual stimuli (and unisensory distractors in the target modality) were presented. The target modality was either alternated on a block-by-block basis (i.e., was predictable; Experiment 1) or else was varied randomly on a trial-by-trial basis (i.e., it was unpredictable; Experiment 2). With this arrangement in place, either a unisensory or a multisensory attentional set was induced as the features of either one or two modalities had to be responded to within an experimental block of trials. In Experiment 3, this modulation was induced within one experiment. Moreover, this experiment included multisensory distractors comprising either both target modalities (audiovisual distractors, as in Experiments 1 and 2) or the current target modality and one neutral modality, namely touch. With these experiments, our aim was to investigate whether attentional set determined, and can be seen as a further moderator of, multisensory distractor processing under conditions of multisensory selection.

To foreshadow the results, the feature of the multisensory distractor that was presented in the same modality as the target (as well as the unisensory distractors) significantly influenced target information processing in both experiments, regardless of the participants' attentional set. The processing of the crossmodal feature of the multisensory distractor, however, was modulated by attentional set. When the target modality was blocked and therefore entirely predictable (Experiment 1), the crossmodal feature did not elicit a significant compatibility effect. By contrast, in Experiment 2, when the modality of the target varied unpredictably on a trial-by-trial basis, the crossmodal feature of the multisensory distractor interfered with target processing. In Experiment 3, the results demonstrated the same pattern of results in a within-participants experimental design. Furthermore, the multisensory distractor feature presented in a neutral modality (touch) failed to interfere with target processing under any condition. This supports the idea that attentional set, in contrast to a general interference of irrelevant input, modulated the impact of the crossmodal feature in the alternating target modality. Within the concept of attentional set, irrelevant input, which is never presented as a target within the flanker task, should not be part of the attentional set, regardless of the predictability of the target modality. Taken together, the results of the three experiments reported in the present study therefore demonstrate that the manipulation of attentional set (i.e., unisensory vs. multisensory; see Fig. 1b) moderated the processing of the

Fig. 1 **a** Bird's-eye view of the experimental setup used in Experiments 1 and 2, and a sample timeline with a congruent-compatible multisensory distractor to a high-pitched auditory target. Note that an LED was affixed to the front of each loudspeaker. **b** Attentional set and feature match of the multisensory distractor with the target within a block. **c** Mean

compatibility effect (RT, in ms) of the crossmodal distractor feature (incompatible minus compatible trials) in the multisensory distractor trials in Experiments 1 and 2. Error bars indicate the standard error of the difference

to-be-ignored multisensory distractor. When the attentional set concerned only the features of a single modality, irrelevant information that was presented in a different sensory modality did not interfere with target processing. However, when the participants' attentional set included features from two modalities, the crossmodal distractor feature presented in the second modality interfered.

Experiment 1

Method

Participants In the literature, and based on our own previous laboratory research, flanker effects typically elicit medium effect sizes (irrespective of the sensory modality in which the stimuli are presented; e.g., Wesslein, Spence, & Frings, 2014). Hence, to ensure that we at least optimized the chances

of observing a unisensory flanker interference effect, we planned our study with a medium effect size in mind (d_z of around .5). Given an alpha-level of .05 and a desired power of at least $1 - \beta > .80$, we aimed for a minimum of 27 participants (power analyses were conducted with GPOWER 3.1.7; Erdfelder, Faul, & Buchner, 1996). In total, 33 students from the University of Trier took part in this study in return for course credit. Three of the participants were excluded from the study due to extremely high error rates (these participants were outliers with respect to the error rates of the sample; $M = 9\%$, $SD = 4\%$, outliers with $> 28\%$ errors). In the end, a total of 30 participants (six male) ages 18 to 46 years ($M = 23$ years, $SD = 5$ years) were used for analyses. All of the participants had normal or corrected-to-normal vision and audition.¹

¹ Note that exactly the same pattern of results was obtained in all of the analyses without excluding the participants.

Apparatus and materials The participants were seated in a dark soundproofed chamber in front of a 24-in. CRT computer screen with a refresh rate of 60 Hz. The distance to the screen (70 cm) was kept constant by means of a chin rest with forehead support. The instructions were presented in grey letters against a black background in the middle of the screen. The fixation cross was also grey. Three tones (low, 440 Hz; medium, 880 Hz; and high, 1760 Hz) and three different colors (red, HSV: 0, 100, 12.55; green, HSV: 120, 100, 12.55; and blue, HSV: 240, 100, 12.55), were used as unisensory target and distractor stimuli. Combinations of auditory and visual stimuli were used as the multisensory distractor. These stimuli were presented from two loudspeakers (Logitech S-120 2.0 OEM, black) and two LEDs were affixed to the lower part of the speakers that were standing laterally on each side of the computer screen. There was a distance of 80 cm between the loudspeakers (corresponding to a visual angle of 60°). See Fig. 1a for a bird's-eye view of the experimental setup. Responses were made with foot pedals using Chronos® (<https://pstnet.com/products/chronos/>). The experiment was conducted with E-Prime (Version 2.0) software, and the data were analyzed with IBM SPSS Statistics (Version 23).

Design The experiment consisted of a $2 \times 2 \times 3$ repeated-measures experimental design with the variables target modality (auditory and visual), distractor condition (multisensory and unisensory), and congruency (congruent, neutral, and incongruent). In the multisensory condition, an audiovisual feature combination was presented that comprised one intramodal and one crossmodal feature defined relative to the target. The intramodal feature was presented in the same modality as the target and could either be congruent, neutral, or incongruent with respect to the target identity and response. The crossmodal feature compromised the other modality and could be compatible, neutral, or incompatible with respect to the target response. The intramodal and crossmodal features of the multisensory distractors were varied orthogonally with respect to each other. For the comparison of unisensory and multisensory trials, the congruency of the multisensory distractor describes the relationship between the target and the intramodal feature.

As the main focus was on the difference between the two multisensory distractor features, those trials with multisensory distractors were analyzed separately in a 3×3 repeated-measures design with the following variables: Intramodal feature congruency (congruent, neutral, and incongruent), and crossmodal feature compatibility (compatible, neutral, and incompatible). Reaction times (RT) and error rates were used as the dependent variables. Note that we left target modality out of the main analyses to simplify interpretation, but report the full analyses in the [Appendix](#). Furthermore, we also report the comparison between multisensory and unisensory trials in the [Appendix](#).

Procedure The participants were tested individually in a soundproofed and dimly illuminated experimental chamber. They sat 70 cm from the display with their head resting on a chin rest. They were instructed to fixate a fixation cross presented in the middle of the screen throughout the experimental blocks. The participants placed their feet on foot pedals. Their task involved classifying the identity of the target by means of foot pedal responses. A left foot-pedal response was required for a high-pitched tone or a red light and a right foot-pedal response for a low-pitched tone or a blue light. The participants were instructed to respond to these target stimuli presented from either side as rapidly and accurately as possible while trying their best to ignore the distractor stimuli presented from the other box. The target modality (visual vs. auditory) was alternated blockwise over eight blocks with 72 samples each (576 trials in total, 288 for each modality), in order to obtain a response assignment for both the auditory and visual stimuli that lasted throughout the whole experiment. The starting modality (visual vs. auditory first) was counterbalanced across participants.

The unisensory distractor was presented in the same modality as the target. It was drawn from a pool of three unisensory visual and three auditory features: one response-neutral feature and the two features that were used as target features (e.g., red and blue color) in each modality. A green light and a medium tone served as neutral features since they were not mapped on to any of the responses. The distractor could either be congruent, neutral, or incongruent with respect to the target. In the case of the multisensory distractor, the intramodal feature (in the target modality) could either be congruent, neutral, or incongruent with respect to the identity of the target (and hence the response required), while the crossmodal feature (in the nontarget modality) could be compatible, neutral, or incompatible with respect to the target response. The intramodal and crossmodal features of the multisensory distractors were varied orthogonally with respect to each other. The nine multisensory distractor conditions accounted for an equal number of trials (32 throughout the whole experiment). Half of all trials were conducted with multisensory distractors, while the remainder had unisensory distractors (96 for each congruency condition). The target (and distractor) side (left, right) was manipulated on a within-participants basis and changed in the middle of the experimental phase of the study (i.e., after four blocks). The order (target on the left/right first) was counterbalanced across participants. Besides, participants were offered a short break with every change of target modality.

The exact stimulus sequence in each trial (see Fig. 1a) was as follows: A fixation cross was presented in the middle of the screen indicating the start of the trial followed by the onset of the distractor after 250 ms. The target was presented 70 ms after the onset of the distractor. Both stimuli were presented for 1,000 ms and the participants had to respond within

1,500 ms of the onset of the distractor. Feedback was only given when the participant made the incorrect response, or else failed to respond altogether. Feedback was displayed for 500 ms. The response-stimulus interval (RSI) was set at 500 ms.

Instructions before the experiment were provided on the screen. The participants were encouraged to ask the experimenter (who was in the control room next to the investigation room during the entire session), if anything was unclear. The experiment started with a learning phase in which the stimulus-response mapping was presented within eight trials. The participants then practiced the task in each modality for 36 trials, respectively. The following experimental phase started with the modality in which they had practiced first, which was counterbalanced across participants. The experiment lasted for about 40 minutes.

Results

For the calculation of the RT, only those trials in which the participants responded correctly to the target were considered. All of the trials in which the RT was shorter than 200 ms, as well as all those trials with an RT that was 1.5 interquartile ranges above the third quartile of each participant's individual RT distribution (Tukey, 1977), were excluded from the data analysis.² In total, 12.4% of the trials were excluded from the analysis of the data due to these restrictions. Table 1 highlights the mean RT and error data for the multisensory distractor trials comparing intramodal and crossmodal feature. Note that the comparison between multisensory and unisensory distractor trials is reported in the [Appendix](#).

RTs As we were particularly interested in the effect of the crossmodal multisensory distractor feature, we only analyzed the multisensory distractor trials in this analysis. A 3 (intramodal feature congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (crossmodal feature compatibility: compatible vs. neutral vs. incompatible) multivariate analysis of variance (MANOVA), with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of intramodal feature congruency, $F(2, 28) = 80.48$, $p < .001$, $\eta_p^2 = .85$, with participants responding fastest when the relevant feature was congruent and slowest when it was incongruent, as expected. Individual contrasts revealed that the difference between congruent and neutral, $t(29) = 6.2$, $p < .001$, as well as the difference between neutral and incongruent,

Table 1 Mean RT (in ms; error rates in %, in parentheses) as a function of intramodal feature congruency and crossmodal feature compatibility in Experiment 1

Crossmodal feature	Intramodal feature			Mean
	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	
<i>Compatible</i>	593 (4.1)	622 (6.7)	645 (17.2)	620 (9.3)
<i>Neutral</i>	598 (5.0)	620 (7.5)	634 (15.3)	617 (9.3)
<i>Incompatible</i>	604 (4.0)	616 (6.7)	646 (15.8)	622 (8.8)
<i>Mean</i>	598 (4.3)	619 (6.9)	642 (16.1)	

$t(29) = 6.5$, $p < .001$, were significant. There was no main effect of crossmodal feature compatibility, $F(2, 28) = 1.11$, $p = .343$, $\eta_p^2 = .07$, and no interaction between intramodal feature congruency and crossmodal feature compatibility, $F(4, 26) = 2.09$, $p = .111$, $\eta_p^2 = .24$. These results suggest that there was no compatibility effect from the irrelevant response feature (see Fig. 1c).

Errors The same 3×3 MANOVA was conducted on the error data and revealed a significant main effect of intramodal feature congruency as well, $F(2, 28) = 29.09$, $p < .001$, $\eta_p^2 = .68$. This result indicates that participants made fewer errors when the intramodal feature was congruent with the target than when it was incongruent. Once again, neither crossmodal feature compatibility, $F(2, 28) = .234$, $p = .793$, $\eta_p^2 = .02$, nor the interaction between the two variables, $F(2, 28) = 1.01$, $p = .419$, $\eta_p^2 = .14$, was significant.

Discussion

In Experiment 1, we used multisensory (and unisensory) distractors in a flanker task with unisensory auditory and visual targets in order to investigate the characteristics of multisensory distractor processing. The distractor feature in the target modality had perceptual overlap while the crossmodal feature in the nontarget modality had response overlap only. Critically, a unisensory attentional set was induced by varying the target modality within blocks of experimental trials and thus, focusing attention on only one modality at a time. With this unisensory attentional set, the intramodal feature of the multisensory distractor elicited an interference effect while the crossmodal feature did not. There was no interaction between the effects of the distractor features. These results suggest that irrelevant information does not interfere with participants' information processing if it is not part of the attentional set, consistent with previous research on attentional capture (e.g., Mast & Frings, 2014). An additional crossmodal feature varying in response compatibility revealed no influence on target processing when presented in a modality that is currently task irrelevant.

² The third quartile of each participant's individual RT distribution reflects the RT value, under which lie 75% of all the RT values for that specific participant. The interquartile range is the range from the 25% to the 75% marker. An RT value that is 1.5 interquartile ranges above the 75% marker is considered as an outlier and is therefore not included into the RT analysis.

To support the assumption that the crossmodal feature shows no effect because of the participants' unisensory attentional set, the attentional set was changed in Experiment 2. The target modality was now manipulated on a trial-by-trial basis, thus making it unpredictable, and hence meaning that the participants presumably had to maintain a multisensory attentional set throughout the course of the experiment. When the modality is not predictable, participants have to include two feature dimensions in their current attentional set. Thus, attention spreads to both modalities. Accordingly, the crossmodal feature might exert a significant influence over participants' performance in this setting.

Experiment 2

Method

Participants Thirty-three students from the University of Trier took part in the experiment in return for course credit. Two of the students were excluded due to extremely high error rates (these participants were outliers with respect to the errors relative to the rest of the sample; $M = 13\%$, $SD = 12\%$, outliers with $> 48\%$ errors). A total of 31 participants (six male) ages 18 to 32 years ($M = 23$ years, $SD = 3$ years) were used for the data analyses. All of the participants had normal or corrected-to-normal vision and audition.³

Apparatus, materials, design, and procedure The apparatus, materials, design, and procedure were exactly the same as in Experiment 1, except for one major change. The target modality was now varied within trials instead of a blockwise manipulation, as in Experiment 1, in order to induce a multisensory attentional set, in which the participants had to include the features of both modalities.

Results

For the calculation of the RT, only those trials in which the participants responded correctly to the target were considered. All of the trials in which the RT was shorter than 200 ms, as well as those trials with an RT that was 1.5 interquartile ranges above the third quartile of each participant's individual RT distribution (Tukey, 1977), were excluded from the data analysis. In total, 8.4% of the trials were excluded from the analysis of the data due to these restrictions. Table 2 depicts the mean RT and error data for the multisensory distractor trials. (Again, the comparison between multisensory and unisensory distractor trials is reported in the [Appendix](#).)

Table 2 Mean RT (in ms; error rates in % in parentheses) as a function of intramodal feature congruency and crossmodal feature compatibility in Experiment 2

Crossmodal feature	Intramodal feature			Mean
	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	
<i>Compatible</i>	729 (7.4)	754 (11.4)	766 (14.5)	750 (11.4)
<i>Neutral</i>	745 (10.0)	770 (15.2)	781 (16.6)	766 (14.0)
<i>Incompatible</i>	756 (16.2)	766 (19.1)	773 (18.0)	765 (17.7)
<i>Mean</i>	743 (11.5)	764 (15.2)	773 (16.4)	

RT For the multisensory distractor trials, a 3 (intramodal feature congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (crossmodal feature compatibility: compatible vs. neutral vs. incompatible) MANOVA, with Pillai's trace as the criterion, was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of intramodal feature congruency, $F(2, 29) = 21.72$, $p < .001$, $\eta_p^2 = .60$; that is, participants responded fastest when the feature was congruent with the target and slowest when it was incongruent. Individual contrasts once again revealed significant differences between the congruent and neutral trials, $t(30) = 5.1$, $p < .001$, as well as between the neutral and incongruent trials, $t(30) = 2.5$, $p = .019$. Critically, the main effect of crossmodal feature compatibility was significant as well, $F(2, 29) = 4.90$, $p = .015$, $\eta_p^2 = .25$, thus suggesting that this feature influenced participants' performance (see Fig. 1c). Individual contrasts revealed that crossmodal compatible trials significantly differed from neutral trials, $t(30) = 3.0$, $p = .006$, while the neutral trials and incompatible trials did not differ significantly, $t(30) = -.138$, $p = .891$. The interaction between intramodal feature congruency and crossmodal feature compatibility was not significant, $F(2, 29) = 1.21$, $p = .331$, $\eta_p^2 = .15$, thus suggesting that the compatibility effect of the crossmodal feature was independent of the effect of the intramodal feature.

Errors The same 3×3 MANOVA was conducted on the error data and revealed an effect of intramodal feature congruency, $F(2, 29) = 5.67$, $p = .008$, $\eta_p^2 = .28$, as well as crossmodal feature compatibility, $F(2, 29) = 3.91$, $p = .031$, $\eta_p^2 = .21$. The interaction between the two variables was not significant, $F(4, 27) = 2.26$, $p = .088$, $\eta_p^2 = .25$.

Between-experiments comparison We further compared the results of the two experiments in order to support the claim that a change of the attentional set led to a significantly different pattern of information processing. A 3 (intramodal feature congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (crossmodal feature compatibility: compatible vs. neutral vs. incompatible) $\times$ 2 (experiment: 1 vs. 2) MANOVA, with Pillai's trace as criterion, and mean RT as the dependent

³ Once again, all of the analyses were also conducted with the full sample ($N = 33$) showing exactly the same pattern of results as the analyses without the excluded participants.

variable, was conducted on the multisensory trials. The analysis revealed a significant interaction between crossmodal feature compatibility and Experiment, $F(2, 58) = 4.86$, $p = .011$, $\eta_p^2 = .14$, indicating that the processing of the crossmodal feature differed between experiments.⁴ Additionally, there was a significant interaction between intramodal feature congruency and experiment, $F(2, 58) = 3.36$, $p = .041$, $\eta_p^2 = .10$, indicating that the congruency effect of the intramodal feature was larger in Experiment 1 than in Experiment 2. The interaction between the intramodal and crossmodal feature was not further modulated by Experiment ($p = .426$).

Sequence analysis To distinguish priming and expectancy-driven effects, we further analyzed if the modality of the previous trial affects processing. Therefore, a 3 (intramodal feature congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (crossmodal feature compatibility: compatible vs. neutral vs. incompatible) $\times$ 2 (previous modality: same vs. different) MANOVA with Pillai's trace as criterion and mean RT as the dependent variable was conducted on the multisensory trials. One participant was dropped from the analysis due to a missing value in one condition. There was a main effect of previous modality, $F(1, 29) = 43.18$, $p < .001$, $\eta_p^2 = .60$, with participants responding more rapidly when the modality of the previous trial was the same as in the current trial than when it was different. There were no further interactions of the other variables with previous modality (all $ps > .108$).

Discussion

In Experiment 2, we again conducted a flanker task with unisensory targets and multisensory distractors. In contrast to Experiment 1, though, a trial-by-trial variation of the target modality was used in order to induce a multisensory attentional set. Intriguingly, this experimental manipulation exerted a significant influence over performance. Next to the congruency effect of the intramodal distractor feature in the target modality, there was now also a significant compatibility effect of the crossmodal distractor feature in the nontarget modality. As in Experiment 1, there was no interaction between the distractor features. The participants were no longer able to focus their attention exclusively on just a single sensory modality, and, hence, the feature in the other modality was processed up to the response level and interfered with participants' performance. These findings support the idea that attentional set modulates multisensory distractor processing. When attentional set is unisensory (including features from one modality), irrelevant information in a different sensory modality (that is not currently part of the attentional set) does not interfere

with target processing. By contrast, the same information does interfere when the participants' attentional set is multisensory and features from two different sensory modalities are task relevant.

To further verify the idea that only features that are part of the attentional set are able to produce an interference effect in multisensory selection, we conducted a third and final experiment in which stimuli from a third modality were used as multisensory distractor features, but not as target features. Accordingly, we adopted tactile stimuli (weak, medium, and strong vibrations) that were presented from the same object as the auditory and visual stimuli. Furthermore, they were mapped to the same responses as the visual and auditory stimuli. This was achieved by having the participants respond to these stimuli in individual blocks of trials without distractors. Critically, tactile stimuli never appeared as targets in the experimental trials with multisensory distractors but could appear as the crossmodal distractor feature instead of the alternating modality (audition when the target was visual).

Experiment 3

Method

Participants Thirty-two students from the University of Trier took part in the experiment in return for course credit or were paid 8€ (~\$9). Six students were excluded due to extremely high error rates. Note that these participants were outliers with respect to the errors relative to the rest of the sample ($M = 11\%$, $SD = 5\%$, outliers with $>30\%$ errors). The data from a total of 26 participants (eight male) aged between the ages of 18 and 32 years ($M = 24$ years, $SD = 3$ years) were used for the data analyses. All of the participants had normal or corrected-to-normal vision and audition.

Apparatus and materials The experimental setup and materials were exactly the same as in Experiment 1 and 2, except for a few changes. Instead of loudspeaker boxes, two custom-made multisensory cubes with two LEDs at the front, a loudspeaker at the back, and two vibration tactors on top and bottom (Model C-2, Engineering Acoustic, Inc.) were used for the presentation of the stimuli (see Merz, Jensen, Spence, & Frings, 2019, for a detailed description of the cubes). This was due to adopting tactile stimuli into the experiment. The upper LEDs as well as the upper tactors were used for the presentation of the visual and tactile stimuli. Tactile stimuli were presented at one of three different vibration intensities (strong: about 200 μm peak-to-peak amplitude; medium: about 56 μm peak-to-peak amplitude; weak: about 15 μm peak-to-peak amplitude, frequency of about 270 Hz), of which one (medium) was used as neutral dimension without a response. To ensure that participants did not judge the identity

⁴ In addition, we repeated the comparison of compatibility effects between Experiments 1 and 2 with RT-adjusted effects as to account for differences in mean RT. The difference was still significant, $t(59) = 2.02$, $p = .042$.

of the vibrations by their auditory, but rather by their tactile impressions, brown noise was presented from two loudspeaker boxes (Wavemaster MX3+ BT-2.1 Stereo) that were standing in the background, next to the computer screen. Brown noise contains low frequencies and was used as the auditory stimuli can still be perceived and classified while the sound of the tactors is absorbed. As participants had to touch the upper tactor of the cube with their index finger, the cubes had to stand closer to the participants compared to Experiments 1 and 2. Accordingly, the distance was individually varied between participants, while the visual angle of 60° was kept constant for all participants.

Design The experiment consisted of a $2 \times 2 \times 2 \times 3 \times 3$ repeated-measures experimental design with the variables target modality (visual vs. auditory), predictability (of the target modality; predictable vs. unpredictable), crossmodal feature identity (crossmodal audiovisual vs. crossmodal tactile), intramodal feature congruency (congruent vs. neutral vs. incongruent), and crossmodal feature compatibility (compatible vs. neutral vs. incompatible). In contrast to Experiments 1 and 2, unisensory distractors were not presented in order to reduce the complexity of the experiment for the participants. The multisensory distractor always compromised an intramodal feature and a crossmodal feature (i.e., there were no trials with two crossmodal features, e.g., visual-tactile distractors for an auditory target). Crossmodal audiovisual distractor trials for example compromise trials with a visual target and an audiovisual distractor while crossmodal tactile trials compromise trials with a visual target and visual-tactile distractor (see Fig. 2a). To further reduce complexity and ease the interpretability of the design, the neutral dimension as well as the target modality were left out of the main analyses.⁵ Accordingly, the reduced design was a $2 \times 2 \times 2$ repeated-measures experimental design with the variables predictability (of the target modality; predictable vs. unpredictable), crossmodal feature identity (crossmodal audiovisual vs. crossmodal tactile), intramodal feature congruency (congruent vs. incongruent), and crossmodal feature compatibility (compatible vs. incompatible). RTs and error rates were used as the dependent variables.

Procedure The participants were tested individually in a soundproofed and dimly illuminated experimental chamber. They sat 70 cm from the display with their head resting on a chin rest. They were instructed to fixate a fixation cross presented in the middle of the screen throughout the experimental blocks. The participants placed their feet on foot pedals. Their task was to classify the identity of the target by means of foot-pedal responses. A left foot-pedal response was required for

a high-pitched tone or a red light and a right foot-pedal response for a low-pitched tone or a blue light. The predictability of the target modality (visual vs. auditory) was manipulated within blocks. A block-wise (predictable) alternation was presented in eight blocks with 36 samples each (288 trials, 144 for each modality), while a trial-by-trial (unpredictable) alternation was presented in four blocks with 72 samples each (288 trials). Predictability alternated every 144 trials (four predictable blocks or two unpredictable blocks). In total, participants completed 576 experimental trials. In order to adopt a neutral modality (i.e., a modality that was never used as target modality in the experimental trials) with a response assignment, six tactile target blocks with eight trials each (48 trials) were mixed in between the experimental blocks (eight trials after every two blocks). In these blocks, tactile stimuli had to be classified with the same foot-pedal responses (left vs. right) as the auditory and visual targets. The experimental phase always started with a tactile block. Whether participants then continued with predictable vs. unpredictable trials was counterbalanced across participants. The predictable trials always started with an auditory block in the first half of the experiment and with a visual block in the second half of the experiment. In the training phase before the experiment, however, the starting modality (visual vs. auditory) was counterbalanced across participants, while the touch was always presented as the third modality.

In contrast to the design of Experiments 1 and 2, unisensory distractors were left out in this experiment to reduce the length and complexity of the study. The multisensory distractor always consisted of a combination of an intramodal feature (i.e., a feature in the target modality) and a crossmodal feature (i.e., a feature in a nontarget modality). The intramodal feature could be congruent, neutral, or incongruent with respect to the target identity and response while the crossmodal feature could be compatible, neutral, or incompatible with respect to the target response. The crossmodal feature further could be in the alternating target modality (crossmodal audiovisual) or in a neutral modality (crossmodal tactile; see Fig. 2a). The intramodal and crossmodal features of the multisensory distractors were varied orthogonally with respect to each other. This resulted in 18 multisensory distractor conditions (nine crossmodal audiovisual and nine crossmodal tactile). As they were presented in predictable and unpredictable blocks, a total of 36 conditions were obtained that accounted for an equal number of trials (16 throughout the experimental blocks). The target (and distractor) side (left, right) was manipulated on a within-participants basis and changed in the middle of the experimental phase (after 6 experimental blocks plus three tactile blocks). The order (target on the left/right first) was counterbalanced across participants. Besides,

⁵ The full analyses are reported in the [Appendix](#).

Fig. 2 **a** Distractor conditions in Experiment 3 for a visual target as an example. **b** Mean compatibility effect (RT, in ms) of the crossmodal distractor feature (incompatible minus compatible trials) as a function of

crossmodal feature identity in Experiment 3. Error bars indicate the standard error of the difference

participants were offered a short break with every change of predictability or modality (when it was blocked).

The stimulus sequence in each trial was the exactly the same as in the previous two experiments (see Fig. 1a). Instructions before the experiment were provided on the screen. The participants were instructed to respond to the target stimuli presented from the target cube as rapidly and accurately as possible while trying their best to ignore the distractor stimuli presented from the other cube. They were encouraged to ask the experimenter (who was in the control room next to the experimental room during the entire experiment), if anything was unclear. The experiment started with a learning phase in which the stimulus-response mapping for one modality (visual vs. auditory first) was presented within eight trials. The participants then practiced the task in this modality for 16 trials, before they learned the mapping of the other two modalities. After that, a training phase with auditory and visual targets and multisensory distractors (equating to an unpredictable experimental block, except that feedback was provided for correct responses as well) was run with 36 trials. These 36 trials were selected randomly from the 72 possible trial combinations. The experiment lasted for about 60 minutes.

Results

For the calculation of the RT, only those trials in which the participants responded correctly to the target were considered. All of the trials in which the RT was shorter than 200 ms, as well as those trials with an RT that was 1.5 interquartile ranges above the third quartile of each participant's individual RT distribution (Tukey, 1977), were excluded from the data analysis. In total, 8.6% of the trials were excluded from the analysis of the data

due to these restrictions. Table 3 depicts the mean RT and error data.

RT A 2 (predictability: predictable vs. unpredictable target modality) $\times$ 2 (crossmodal feature identity: crossmodal audiovisual vs. crossmodal tactile) $\times$ 2 (intramodal feature congruency: congruent vs. neutral vs. incongruent) $\times$ 2 (crossmodal feature compatibility: compatible vs. neutral vs. incompatible) MANOVA with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of predictability, $F(1, 25) = 75.0$, $p < .001$, $\eta_p^2 = .75$, with participants responding more rapidly when the target modality was predictable than when it was unpredictable. There was a main effect of intramodal feature congruency, $F(1, 25) = 41.47$, $p < .001$, $\eta_p^2 = .62$, and crossmodal feature compatibility, $F(1, 25) = 7.70$, $p = .01$, $\eta_p^2 = .24$, in the expected direction (participants responded fastest when the feature was congruent/compatible with the target and slowest when it was incongruent/incompatible). The main effect of crossmodal feature identity was not significant ($F < 1$), indicating that participants were equally fast in responding to crossmodal audiovisual and crossmodal tactile distractors. Critically, however, there was tentative evidence for an interaction of crossmodal feature identity and crossmodal feature compatibility, $F(1, 25) = 3.24$, $p = .084$, $\eta_p^2 = .12$.⁶ Further analyses

⁶ Note that an interaction test in a within-participants design with a numerator df of 1 is formally equivalent to a one-sample t test, with $t = \sqrt{F}$ and $p(F) = p(t)$ (two-tailed). In order to transform the interaction result of the F test in a t test, we have to calculate the critical difference variable (i.e., compatibility effects), and test it against zero. Given our specific predictions, a one-tailed test is permissible (see Maxwell & Delaney, 2004, p. 164), this t test revealed a significant result, $t(25) = 1.80$, $p = .042$, $\eta_p^2 = .12$ (one-tailed).

Table 3 Mean RT (in ms; error rates in % in parentheses) as a function of crossmodal feature identity, intramodal feature congruency, and crossmodal feature compatibility in Experiment 3

Crossmodal audiovisual	Intramodal feature			Mean
Crossmodal feature	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	
<i>Compatible</i>	753 (6.7)	782 (11.2)	794 (15.3)	776 (11.1)
<i>Neutral</i>	753 (6.4)	789 (11.9)	803 (16.0)	782 (11.4)
<i>Incompatible</i>	766 (9.4)	784 (15.0)	811 (14.4)	787 (13.0)
<i>Mean</i>	757 (7.5)	785 (12.7)	803 (15.2)	
Crossmodal tactile				
<i>Compatible</i>	763 (8.1)	787 (12.9)	794 (15.4)	782 (12.1)
<i>Neutral</i>	759 (7.9)	778 (11.4)	789 (13.6)	775 (11.0)
<i>Incompatible</i>	767 (7.0)	781 (12.0)	795 (14.1)	781 (11.0)
<i>Mean</i>	763 (7.7)	782 (12.7)	793 (14.3)	

revealed that the compatibility effect of the crossmodal distractor feature was significant for crossmodal audiovisual distractors, $F(1, 25) = 11.26$, $p = .003$, $\eta_p^2 = .31$, but not for crossmodal tactile distractors ($F < 1$; see Fig. 2b). This effect was not further modulated by predictability or intramodal feature congruency (all p s $> .309$). Further, there was a significant interaction of predictability and crossmodal feature compatibility, $F(1, 25) = 4.34$, $p = .048$, $\eta_p^2 = .15$, indicating that there was no effect of the crossmodal feature in the predictable blocks while there was an effect in the unpredictable blocks. Besides that, there was a significant interaction of predictability and intramodal feature congruency, $F(1, 25) = 24.22$, $p < .001$, $\eta_p^2 = .49$, indicating a larger congruency effect in the predictable blocks than in the unpredictable blocks. Furthermore, a significant interaction between predictability and crossmodal feature identity evolved, $F(1, 25) = 7.00$, $p = .014$, $\eta_p^2 = .22$. In predictable blocks, the participants responded more rapidly when the distractor was crossmodal audiovisual than when it was crossmodal tactile, whereas it was the other way around in the unpredictable blocks. None of the other effects were significant (all p s $> .374$).

Errors For the error rates, the same $2 \times 2 \times 2 \times 2$ MANOVA as for the RT data was conducted, and no contradictory effects were obtained. The analysis revealed a significant main effect of predictability in the expected direction, $F(1, 25) = 27.71$, $p < .001$, $\eta_p^2 = .53$, and a significant main effect of intramodal feature congruency in the expected direction, $F(1, 25) = 41.01$, $p < .001$, $\eta_p^2 = .62$. None of the other effect were significant (all p s $> .140$).

Audiovisual distractor trials Additionally, the audiovisual distractor trials were again analyzed separately to replicate

the between-comparison analysis of Experiments 1 and 2 within a repeated-measures analysis.

RT A 2 (predictability: predictable vs. unpredictable target modality) $\times$ 2 (intramodal feature congruency: congruent vs. neutral vs. incongruent) $\times$ 2 (crossmodal feature compatibility: compatible vs. neutral vs. incompatible) MANOVA, with Pillai's trace as the criterion, was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of predictability, $F(1, 25) = 67.73$, $p < .001$, $\eta_p^2 = .73$, with those trials where the target modality was predictable being responded to faster than unpredictable trials. Further, there were significant main effects of intramodal feature congruency, $F(1, 25) = 39.73$, $p < .001$, $\eta_p^2 = .61$, and crossmodal feature compatibility, $F(1, 25) = 13.42$, $p = .001$, $\eta_p^2 = .35$, with participants responding more rapidly on congruent/compatible trials than on incongruent/incompatible trials. Critically, the Predictability $\times$ Crossmodal Feature Compatibility interaction reached marginal one-tailed significance, $F(1, 25) = 2.70$, $p = .06$ (one-tailed),⁷ $\eta_p^2 = .10$, thus suggesting that the compatibility effect of the crossmodal feature was only present when the target modality was unpredictable, $F(1, 25) = 8.98$, $p = .006$, $\eta_p^2 = .26$, compared with when it was predictable, $F(1, 25) = 0.65$, $p = .428$, $\eta_p^2 = .03$. Additionally, there was an interaction between predictability and intramodal feature congruency, $F(1,$

⁷ Once again, there was at least tentative evidence for the interaction of predictability and crossmodal feature compatibility, $F(1, 25) = 2.70$, $p = .112$, $\eta_p^2 = .10$. The interaction test in a within-participants design with a numerator df of 1 is formally equivalent to a one-sample t test, with $t = \sqrt{F}$ and $p(F) = p(t)$ (two-tailed). In order to transform the interaction result of the F test in a t test, we have to calculate the critical difference variable (i.e., compatibility effects), and test it against zero. Given our specific predictions, a one-tailed test is permissible (see Maxwell & Delaney, 2004, p. 164), $t(25) = 1.64$, $p = .06$, $\eta_p^2 = .12$ (one-tailed).

25) = 15.58, $p = .001$, $\eta_p^2 = .38$, indicating a larger congruency effect for the intramodal feature when the target modality was predictable. None of the other effects were significant (all $ps > .213$).

Errors The same MANOVA as for the RT data was conducted on the error data and revealed significant main effects of predictability, $F(1, 25) = 30.66$, $p < .001$, $\eta_p^2 = .55$, and intramodal feature congruency, $F(1, 25) = 27.79$, $p < .001$, $\eta_p^2 = .53$, and a marginal significant interaction of predictability and intramodal feature congruency, $F(1, 25) = 2.77$, $p = .10$, $\eta_p^2 = .10$,⁸ in the expected directions. Further, the three way interaction was significant, $F(1, 25) = 4.47$, $p = .045$, $\eta_p^2 = .15$, indicating that the intramodal and crossmodal feature interacted when the target modality was unpredictable, $F(1, 25) = 4.84$, $p = .037$, $\eta_p^2 = .16$, but not when it was predictable ($F < 1$).

Discussion

In Experiment 3, a flanker task with unisensory auditory and visual targets and multisensory distractors was conducted. The multisensory distractors were either audiovisual or else contained a crossmodal tactile feature (visual tactile, auditory tactile). The idea was that a neutral modality that was not used as the target modality should not produce any interference effects as it was never part of the attentional set that was adopted by participants. Next to this new distractor condition, the predictability of the target modality (blockwise vs. trialwise variation) was manipulated within the experiment. As in the previous experiments, the predictability of the target modality comprised a modulation of the attentional set. Participants can concentrate on the features of (or selectively attend to) one modality when the modality is blocked (unisensory attentional set). Otherwise, they have to keep the features of two modalities in mind when it is varied unpredictably trial by trial (multisensory attentional set). The results showed that the compatibility effect of the crossmodal feature was significantly affected by the modality of that feature. When the crossmodal distractor feature was in the alternating target modality (i.e., auditory/visual), a reliable congruency effect was elicited, while there was no effect when the feature compromised a neutral modality

(i.e., tactile). This was not further modulated by the predictability of the target modality. However, a separate analysis on the audiovisual distractor trials revealed a modulation of the crossmodal compatibility effect by predictability, as expected from the previous experiments. Consequently, the result of the between-experiment comparison of the effects in Experiments 1 and 2 could be replicated and verified. When the target modality was blocked, the crossmodal feature of the audiovisual distractor did not elicit any interference. When the target modality was varied, however, the feature gave rise to a compatibility effect. Accordingly, the results of Experiment 3 further support the idea of the modulation of multisensory distractor processing by attentional set.

General discussion

Across three experiments, we investigated influences of attention on the processing of irrelevant stimuli in multisensory interference tasks. Unisensory auditory and visual targets were presented with multisensory (and unisensory) distractors. Critically, attention was manipulated by inducing different attentional sets. The concept of attentional set goes back to the literature on contingent attentional capture (e.g., Remington & Folk, 2001; see also Hommel et al., 2014, for more recent research on attentional capture) and refers to the set of stimuli or the feature dimensions that are currently responded to, and thus are processed with priority. Here, a modulation of attentional set was implemented by modulating the predictability of the target modality, which significantly influenced multisensory distractor processing. When the target modality was blocked, the intramodal feature of the multisensory distractor produced an interference effect while the crossmodal feature did not. When the target modality was varied, the effects of the two features added up. In this case, the crossmodal distractor feature was processed up to the response level and interfered with target processing. Additionally, a crossmodal distractor feature in a neutral modality (touch) never elicited interference.

In Experiment 1, the modality of the target was kept constant within each block of experimental trials and only varied between successive blocks. This reflects the implementation of a unisensory attentional set as the target modality was predictable and ignoring the features of the alternating modality possible. In this case, there was no interference from the crossmodal feature distractor in the alternating target modality. Thus, participants presumably did not include the features of the other sensory modality into their attentional set (i.e.,

⁸ Once again, there was at least tentative evidence for the interaction of predictability and crossmodal feature compatibility, $F(1, 25) = 2.77$, $p = .108$, $\eta_p^2 = .10$. The interaction test in a within participants design with a numerator df of 1 is formally equivalent to a one-sample t test, with $t = \sqrt{F}$ and $p(F) = p(t)$ (two-tailed). In order to transform the interaction result of the F test in a t test, we have to calculate the critical difference variable (i.e., compatibility effects), and test it against zero. Given our specific predictions, a one-tailed test is permissible (see Maxwell & Delaney, 2004, p. 164), $t(25) = 1.66$, $p = .05$, $\eta_p^2 = .10$ (one-tailed).

they selectively attend to one modality). In contrast, the modality of the target changed unpredictably from trial to trial in Experiment 2, thus presumably inducing an attentional set that was, in some sense, multisensory. This resulted in a reliable compatibility effect of the crossmodal distractor feature, thus suggesting that participants included the features of the two target modalities in the attentional set (i.e., they divided their attention). In Experiment 3, this finding was replicated with in a repeated measures design. Additionally, a new type of multisensory distractors, where the crossmodal feature was presented in a neutral modality (i.e., touch), were introduced in this experiment. Our aim was to verify the idea of attentional set as the features of a neutral modality – that is, a modality that never appeared as a target modality – should not be included in the attentional set. As expected, this distractor features never elicited a measurable interference effect.

Taken together, the results of the three experiments reported here provide a number of important insights. First of all, they are in line with the idea of contingent attentional capture (e.g., Mast & Frings, 2014; Mast, Frings, & Spence, 2014, 2015, 2017; Remington & Folk, 2001), but successfully demonstrate a rather new manipulation of attentional set obtained by varying the predictability of the target modality. Previous research finding significant effects of multisensory distractors that were presented with unisensory targets (Fong et al., 2018; Santangelo, Ho, & Spence, 2008; Santangelo & Spence, 2007) can also be explained in terms of the proposed attentional set approach. In a flanker task with visual targets (Fong et al., 2018), the audiovisual distractor comprised additional auditory input which described the identity of the current visual distractor feature. As such, the auditory feature was redundant with the visual feature and thus, was presumably part of the attentional set. Similarly, in the spatial cueing paradigm used by Santangelo and colleagues, the cueing features from different modalities provided redundant spatial information. In line with our approach, Lunn et al. (2019) did not find attentional capture of redundant multisensory distraction under high load, when it was unrelated to the task, and hence not part of the attentional set.

Most importantly, the present findings support previous research on the influence of attention on multisensory distractor processing with previously unrelated stimuli (Jensen et al., 2019; Merz et al., 2019) and unisensory distractor processing with a combination of feature as stimuli (Lavie, 1997). This study now indicates that other types of attention can also influence the processing of irrelevant information presented in multiple sensory modalities. The manipulation of spatial attention in the way

it was done in our previous studies (i.e., by changing the object in the focus of view) can be considered as an overt way of manipulating attention as the constraint where to look at comes from the experimental arrangement and instructions. The manipulation of attentional set, on the other hand, rather reflects a covert manipulation of attention guided by endogenous strategies (i.e., there was no change in orientation of sensory receptors themselves). Crucially, multisensory distractor processing would seem to be affected by both types of manipulations. Accordingly, the present study extends the knowledge on the relation of attention and the processing of irrelevant input in multisensory selection.

When target modality is modulated on a trial-by-trial basis, one could argue that the priming effects of the previous trial account for the interference of the irrelevant distractor feature (e.g., in Experiment 2; see Spence et al., 2001). In this case, the modality of the irrelevant distractor feature would be primed if it was the same modality as the target in the previous trial. To distinguish whether the effect of the crossmodal distractor feature is modulated top down, by attentional set, or is rather based on passive bottom-up priming, an additional analysis controlling for modality sequence effects was conducted. However, this analysis revealed no modulation of the crossmodal distractor feature interference as a function of the modality of the target on the previous trial. As such, passive priming cannot account for the present findings.

In Experiment 3, an additional multisensory distractor type was added to the experimental design. In addition to audiovisual distractors, multisensory distractors with crossmodal tactile feature were added in order to present a neutral modality as crossmodal distractor feature. As suggested by a friendly reviewer, the idea was to show that a neutral feature dimension, which is never part of the attentional set, could also never elicit a crossmodal interference effect. If so, the proposed modulation of attentional set would be verified to some extent. However, it can be questioned whether the implemented design actually reflects a modulation of attentional set. One could argue that in case of a target modality that is blocked, the other modality is simply suppressed while it is not in the case of a varying modality. Admittedly, the present data cannot account for the distinction between attentional set and modality suppression. However, we do not consider this problematic as these two possibilities are not mutually exclusive. In fact, it is very likely that both mechanisms act together. As such, we do not consider it imperative to resolve this distinction at the present time. In future research, one way to address this issue might to vary the features that belong to the attentional set within one sensory modality. In

particular, there could be one feature belonging to the attentional set and one not belonging in each modality. In this case, the dependence of crossmodal interference on attentional set would be measured without the possibility of modality suppression. Nevertheless, as mentioned, one can never exclude the possibility that both mechanisms account for the present results.

In our previous studies (Jensen et al., 2019; Merz et al., 2019), spatial attention was manipulated and modulated multisensory distractor processing. The features of the distractor were interacting when considered with spatial attention, but were processed more independently when not (see also Navarra, Alsius, Soto-Faraco, & Spence, 2010; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010, for this logic). In the present study, an interaction of the distractor features was not found, indicating that they were processed rather independently from one another. Accordingly, the effects of intramodal and crossmodal features were additive, rather than multiplicative. On the basis of previous studies (Jensen et al., 2019; Merz et al., 2019), an interaction of the distractor features could have been expected as (spatial) attention was shown to modulate the integration process (i.e., the interaction), rather than the individual interference effects. However, the influence of each type of attention might be slightly different. For example, spatial attention was found to influence attentional capture only when the stimuli were part of the attentional set (Huang et al., 2016). In this case, the internal attentional setting acted as a precondition for the influence of spatial attention. Furthermore, there were some important differences concerning the experimental setup between previous studies and the present one, of which the focus of view, and thus the attentional focus would be the most important. In the previous studies, an interaction occurred when the focus was on the distractor. In the present experiments, however, it was always between target and distractor, thus providing a neutral attentional focus. Accordingly, the distractor was never the direct focus of attention and resources directed to the distractor were apparently insufficient to induce integration. One could have thought, that with the use of unisensory targets, the task became easier and provided more spare/available resources for distractor processing. However, even if the task was easier, this was not sufficient for integration when the distractor is not fixated. Another alternative explanation could be that integration is less promoted in this task as a unisensory target does not need integration.

Conclusions

The present study investigated the influence of the attentional set on multisensory distractor processing in a

unisensory target setting. Attentional set was modulated by manipulating the predictability of the target's modality and thus the activation of the corresponding modality-specific response assignments. This manipulation significantly influenced participants' performance. When the target modality was predictable, the distractor feature in the nontarget modality (what is referred to as the crossmodal feature) failed to give rise to any compatibility effect. Participants focused their attention within one block of trials on the features of one modality and excluded the other from their current attentional set. Therefore, a crossmodal distractor feature could not produce measurable interference. However, when the target modality was unpredictable, features of both modalities had to be included into the attentional set. In this case, the crossmodal feature elicited a significant interference effect. This shows, once again, that attention is an important moderator as far as multisensory distractor processing is concerned.

Author note Anne Jensen, Simon Merz, and Christian Frings, University of Trier, Department of Psychology, D-54286, Germany, and Charles Spence, Crossmodal Research Laboratory, Department of Experimental Psychology, Anna Watts Building, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK.

The research reported in this article was supported by a grant from the Deutsche Forschungsgemeinschaft to Christian Frings and Charles Spence (FR 2133/5-3).

Open practices statement The data and codes for all experiments are available at PsychArchives (<https://doi.org/10.23668/psycharchives.2362>).

Appendix

Additional analysis for Experiments 1 and 2

Comparison of multisensory and unisensory distractor trials in Experiments 1 and 2 Note that the analyses of the error data were left out and are only reported when they revealed trends that opposed those reported in the analyses of the RT data. Table 4 depicts the mean RT and error data for both experiments.

Experiment 1 Additionally, multisensory distractor interference was compared to unisensory distractor interference. We looked at the difference between the effect of the intramodal feature of the multisensory distractor and unisensory distractors. A 2 (Distractor Condition: multisensory vs. unisensory) $\times$ 3 (Congruency: congruent vs. neutral vs. incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The analysis revealed a significant

Table 4 Mean RT (in ms; error rates in %, in parentheses) as a function of modality, distractor condition, and congruency in Experiment 1

Congruency	Target modality				Mean
	Visual	Auditory	MS	US	
Experiment 1	MS	US	MS	US	Mean
Congruent	568 (3.1)	561 (4.7)	620 (5.0)	586 (4.7)	583 (4.4)
Neutral	603 (4.2)	585 (5.2)	640 (10.8)	625 (9.5)	611 (7.4)
Incongruent	627 (5.8)	608 (6.8)	672 (25.8)	653 (23.1)	637 (15.4)
Mean	592 (5.0)		633 (13.2)		
Experiment 2					
Congruent	568 (3.1)	561 (4.7)	620 (5.0)	586 (4.7)	583 (4.4)
Neutral	603 (4.2)	585 (5.2)	640 (10.8)	625 (9.5)	611 (7.4)
Incongruent	627 (5.8)	608 (6.8)	672 (25.8)	653 (23.1)	637 (15.4)
Mean	592 (5.0)		633 (13.2)		

Notes. MS = Multisensory distractor; US = Unisensory distractor.

main effect of Distractor Condition, $F(1, 29) = 86.68$, $p < .001$, $\eta_p^2 = .75$, with participants responding more rapidly when the distractor was unisensory than when it was multisensory. There was a significant main effect of Congruency, $F(2, 28) = 173.02$, $p < .001$, $\eta_p^2 = .93$, with participants responding fastest when the distractor was congruent with the target and slowest when it was incongruent. There was no significant interaction between Distractor Condition and Congruency, $F(2, 28) = 2.89$, $p = .072$, $\eta_p^2 = .17$, thus suggesting that congruency effects did not differ between the multisensory and unisensory distractor conditions.

Experiment 2 To compare multisensory and unisensory distractors, a 2 (Distractor Condition: unisensory vs. multisensory) $\times$ 3 (Congruency: congruent vs. neutral vs. incongruent) MANOVA with Pillai's trace as the criterion was conducted, with mean RT as the dependent variable. The MANOVA revealed a significant main effect of Distractor Condition, $F(1, 30) = 46.92$, $p < .001$, $\eta_p^2 = .61$, with participants responding more rapidly when the distractor was unisensory as compared to when it was multisensory. There was a significant main effect of Congruency, $F(2, 29) = 35.01$, $p < .001$, $\eta_p^2 = .71$, with participants responding fastest when the distractor was congruent with the target and slowest when it was incongruent. Critically, the interaction between Distractor Condition and Congruency reached significance, $F(2, 29) = 3.35$, $p = .049$, $\eta_p^2 = .19$, suggesting that the magnitude of the congruency effect for unisensory distractors was larger than for multisensory distractors.

The 2 (Distractor Condition: unisensory vs. multisensory) $\times$ 3 (Congruency: congruent vs. neutral vs. incongruent) $\times$ 2 (Experiment: 1 vs. 2) MANOVA with Pillai's trace as criterion and mean RT as the dependent variable

revealed no significant interaction between Distractor Condition, Congruency, and Experiment, $F(2, 58) = .24$, $p = .784$, $\eta_p^2 = .01$, thus suggesting that the congruency effects of unisensory and multisensory distractors did not differ between the two experiments. However, there was a significant Congruency $\times$ Experiment interaction, $F(2, 58) = 3.41$, $p = .04$, $\eta_p^2 = .11$, suggesting that overall, the congruency effects differed between experiments.

Effects of Modality in Experiments 1 and 2 We further ran all analyses including the factor of Target Modality, and this revealed a number of significant effects. Tables 5 and 6 depict the mean RT and error data, Table 7 the MANOVA for the comparison of intramodal and crossmodal features of the multisensory distractor in Experiments 1 and 2. We left these analyses out of the main text in order to simplify the interpretation of the effects and concentrate on the main message.

Table 5 Mean RT (in milliseconds; error rates in %, in parentheses) as a function of modality, intramodal congruency and crossmodal compatibility in Experiment 1.

Crossmodal feature	Intramodal feature			Mean
	Congruent	Neutral	Incongruent	
Vision				
Compatible	568 (3.1)	600 (4.2)	623 (5.6)	597 (4.3)
Neutral	572 (4.2)	603 (4.2)	622 (5.4)	599 (4.6)
Incompatible	592 (3.1)	603 (4.4)	627 (5.8)	607 (4.4)
Mean	577 (3.5)	602 (4.2)	624 (5.6)	
Audition				
Compatible	620 (5.0)	649 (9.2)	674 (28.8)	648 (14.3)
Neutral	622 (5.8)	640 (10.8)	645 (25.2)	636 (14.0)
Incompatible	615 (4.8)	631 (9.0)	672 (25.8)	639 (13.2)
Mean	619 (5.2)	640 (9.7)	664 (26.6)	

Table 6 Mean RT (in milliseconds; error rates in %, in parentheses) as a function of modality, intramodal congruency and crossmodal compatibility in Experiment 2.

Crossmodal feature	Intramodal feature			Mean
	Congruent	Neutral	Incongruent	
Vision				
Compatible	568 (3.1)	600 (4.2)	623 (5.6)	597 (4.3)
Neutral	572 (4.2)	603 (4.2)	622 (5.4)	599 (4.6)
Incompatible	592 (3.1)	603 (4.4)	627 (5.8)	607 (4.4)
Mean	577 (3.5)	602 (4.2)	624 (5.6)	
Audition				
Compatible	620 (5.0)	649 (9.2)	674 (28.8)	648 (14.3)
Neutral	622 (5.8)	640 (10.8)	645 (25.2)	636 (14.0)
Incompatible	615 (4.8)	631 (9.0)	672 (25.8)	639 (13.2)
Mean	619 (5.2)	640 (9.7)	664 (26.6)	

Table 7 Effects of the MANOVAs on the RT data for the multisensory distractor trials with the variables Target Modality, Intramodal Feature Congruency, and Crossmodal Feature Compatibility, and the

MANOVAs for the comparison for unisensory and multisensory distractor trials with the variables Target Modality, Distractor Condition, and Congruency in Experiments 1 and 2.

Experiment 1

2 (Target Modality: visual vs. auditory) $\times$ 3 (Intramodal Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Crossmodal Feature Compatibility: compatible vs. neutral vs. incompatible).

	<i>F</i>	<i>Degrees of freedom</i>	<i>p</i>	η_p^2
<i>Target Modality</i>	8.29	1, 29	.007	.22
<i>Intramodal Feature (IF)</i>	93.77	2, 28	<.001	.87
<i>Crossmodal Feature (CF)</i>	2.06	2, 28	.147	.13
<i>IF $\times$ Modality</i>	0.12	2, 28	.889	.01
<i>CF $\times$ Modality</i>	5.95	2, 28	.007	.30
<i>IF $\times$ CF</i>	2.79	4, 26	.048	.30
<i>IF $\times$ CF $\times$ Modality</i>	1.46	4, 26	.244	.18

2 (Target Modality: visual vs. auditory) $\times$ 2 (Distractor Condition: multisensory vs. unisensory) $\times$ 3 (Congruency: congruent vs. neutral vs. incongruent)

<i>Target Modality</i>	8.43	1, 29	.007	.23
<i>Distractor Condition (DC)</i>	82.38	1, 29	<.001	.74
<i>Congruency</i>	179.24	2, 28	<.001	.93
<i>DC $\times$ Modality</i>	0.48	1, 29	.494	.02
<i>Congruency $\times$ Modality</i>	.832	2, 28	.446	.06
<i>DC $\times$ Congruency</i>	3.46	2, 28	.045	.20
<i>DC $\times$ Congruency $\times$ Modality</i>	4.49	2, 28	.02	.24

Experiment 2

2 (Target Modality: visual vs. auditory) $\times$ 3 (Intramodal Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Crossmodal Feature Compatibility: compatible vs. neutral vs. incompatible).

	<i>F</i>	<i>Degrees of freedom</i>	<i>p</i>	η_p^2
<i>Target Modality</i>	17.64	1, 30	<.001	.37
<i>Intramodal Feature (IF)</i>	20.81	2, 29	<.001	.59
<i>Crossmodal Feature (CF)</i>	3.76	2, 29	.035	.21
<i>IF $\times$ Modality</i>	0.16	2, 29	.856	.01
<i>CF $\times$ Modality</i>	3.17	2, 29	.057	.18
<i>IF $\times$ CF</i>	1.51	4, 27	.228	.18
<i>IF $\times$ CF $\times$ Modality</i>	0.29	4, 27	.885	.04

2 (Target Modality: visual vs. auditory) $\times$ 2 (Distractor Condition: multisensory vs. unisensory) $\times$ 3 (Congruency: congruent vs. neutral vs. incongruent).

<i>Target Modality</i>	27.15	1, 30	<.001	.48
<i>Distractor Condition (DC)</i>	41.68	1, 30	<.001	.58
<i>Congruency</i>	34.58	2, 29	<.001	.71
<i>DC $\times$ Modality</i>	14.33	1, 30	.001	.32
<i>Congruency $\times$ Modality</i>	0.92	2, 29	.410	.06
<i>DC $\times$ Congruency</i>	3.56	2, 29	.041	.20
<i>DC $\times$ Congruency $\times$ Modality</i>	6.94	2, 29	.003	.32

Notes. Bold italic values indicate two-tailed, italic values one-tailed significance.

Additional Analyses for Experiment 3

In order to try and simplify, we report the main analysis of Experiment 3 with neutral trials included outside the main text. The 2 (Predictability: predictable vs. unpredictable) $\times$ 2 (Crossmodal Feature Identity: crossmodal audiovisual vs. crossmodal tactile) $\times$ 3 (Intramodal Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Crossmodal Feature Compatibility: congruent vs. neutral vs. incongruent) MANOVA with Pillai's trace as criterion and mean RT and error rates as the dependent variables is depicted in Table 8.

Additionally, we also report effects of target modality in a 2 (Target Modality: visual vs. auditory) $\times$ 2 (Predictability: predictable vs. unpredictable) $\times$ 2 (Crossmodal Feature Identity: crossmodal audiovisual vs. crossmodal tactile) $\times$ 3 (Intramodal Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Crossmodal Feature Compatibility: congruent vs. neutral vs. incongruent) MANOVA with Pillai's trace as criterion and mean RT as the dependent variable (see Table 9), as for Experiments 1 and 2. Note that target modality did not modulate our critical interactions. Table 10 depicts the mean RT and error data.

Table 8 Effects of the MANOVAs for the multisensory distractor trials in Experiment 3 with the variables Predictability, Crossmodal Feature Identity, Intramodal Feature Congruency, and Crossmodal Feature Compatibility in the RT and error data.

Experiment 3				
2 (Predictability: predictable vs. unpredictable) × 2 (Crossmodal Feature Identity: crossmodal audiovisual vs. crossmodal tactile) × 3 (Intramodal Feature Congruency: congruent vs. neutral vs. incongruent) × 3 (Crossmodal Feature Compatibility: congruent vs. neutral vs. incongruent).				
RT	<i>F</i>	<i>Degrees of freedom</i>	<i>p</i>	η_p^2
<i>Predictability</i>	85.74	1, 25	< .001	.77
<i>Crossmodal Feature Identity (CFI)</i>	0.74	1, 25	.400	.03
<i>Intramodal Feature (IF)</i>	37.29	2, 24	< .001	.76
<i>Crossmodal Feature (CF)</i>	3.18	2, 24	.060	.21
<i>Predictability × CFI</i>	2.77	1, 25	.109	.10
<i>Predictability × IF</i>	18.27	2, 24	< .001	.60
<i>Predictability × CF</i>	2.84	2, 24	.078	.19
<i>CFI × IF</i>	2.56	2, 24	.098	.18
<i>CFI × CF</i>	6.71	2, 24	.005	.36
<i>IF × CF</i>	0.59	4, 22	.672	.10
<i>Predictability × CFI × IF</i>	1.04	2, 24	.369	.08
<i>Predictability × CFI × CF</i>	0.88	2, 24	.429	.07
<i>Predictability × IF × CF</i>	0.57	4, 22	.684	.10
<i>Predictability × CFI × IF × CF</i>	0.50	4, 22	.739	.08
Errors				
<i>Predictability</i>	28.07	1, 25	< .001	.53
<i>Crossmodal Feature Identity (CFI)</i>	0.65	1, 25	.429	.03
<i>Intramodal Feature (IF)</i>	31.80	2, 24	< .001	.73
<i>Crossmodal Feature (CF)</i>	0.78	2, 24	.469	.06
<i>Predictability × CFI</i>	1.71	1, 25	.203	.06
<i>Predictability × IF</i>	4.50	2, 24	.02	.27
<i>Predictability × CF</i>	0.03	2, 24	.970	.00
<i>CFI × IF</i>	0.63	2, 24	.540	.05
<i>CFI × CF</i>	2.98	2, 24	.070	.20
<i>IF × CF</i>	1.07	4, 22	.396	.16
<i>Predictability × CFI × IF</i>	1.95	2, 24	.164	.14
<i>Predictability × CFI × CF</i>	0.00	2, 24	.996	.00
<i>Predictability × IF × CF</i>	1.67	4, 22	.192	.23
<i>Predictability × CFI × IF × CF</i>	0.85	4, 22	.509	.13

Notes. Bold italic values indicate two-tailed, italic values one-tailed significance.

Table 9 Effects of the MANOVAs for the multisensory distractor trials in Experiment 3 with the variables Target Modality, Predictability, Crossmodal Feature Identity, Intramodal Feature Congruency, and Crossmodal Feature Compatibility in the RT data

Experiment 3

2 (Target Modality: visual vs. auditory) $\times$ 2 (Predictability: predictable vs. unpredictable) $\times$ 2 (Crossmodal Feature Identity: crossmodal audiovisual vs. crossmodal tactile) $\times$ 3 (Intramodal Feature Congruency: congruent vs. neutral vs. incongruent) $\times$ 3 (Crossmodal Feature Compatibility: congruent vs. neutral vs. incongruent).

	<i>F</i>	<i>Df</i>	<i>p</i>	η_p^2
<i>(Target) Modality</i>	71.29	1, 25	< .001	.74
<i>Predictability</i>	88.11	1, 25	< .001	.78
<i>Crossmodal Feature Identity (CFI)</i>	1.00	1, 25	.327	.04
<i>Intramodal Feature (IF)</i>	39.67	2, 24	< .001	.77
<i>Crossmodal Feature (CF)</i>	3.40	2, 24	.05	.22
<i>Modality $\times$ Predictability</i>	0.00	1, 25	.951	.00
<i>Modality $\times$ CFI</i>	1.62	1, 25	.215	.06
<i>Modality $\times$ IF</i>	12.96	2, 24	< .001	.52
<i>Modality $\times$ CF</i>	0.11	2, 24	.900	.01
<i>Predictability $\times$ CFI</i>	3.61	1, 25	.069	.13
<i>Predictability $\times$ IF</i>	14.11	2, 24	< .001	.54
<i>Predictability $\times$ CF</i>	2.48	2, 24	.105	.17
<i>CFI $\times$ IF</i>	2.83	2, 24	.079	.19
<i>CFI $\times$ CF</i>	4.65	2, 24	.02	.28
<i>IF $\times$ CF</i>	0.70	4, 22	.597	.11
<i>Modality $\times$ Predictability $\times$ CFI</i>	2.22	1, 25	.149	.08
<i>Modality $\times$ Predictability $\times$ IF</i>	9.12	2, 24	.001	.43
<i>Modality $\times$ Predictability $\times$ CF</i>	0.07	2, 24	.936	.01
<i>Modality $\times$ CFI $\times$ IF</i>	8.77	2, 24	.001	.42
<i>Modality $\times$ CFI $\times$ CF</i>	1.95	2, 24	.164	.14
<i>Modality $\times$ IF $\times$ CF</i>	4.78	4, 22	.006	.47
<i>Predictability $\times$ CFI $\times$ IF</i>	0.83	2, 24	.450	.06
<i>Predictability $\times$ CFI $\times$ CF</i>	1.56	2, 24	.232	.12
<i>Predictability $\times$ IF $\times$ CF</i>	0.81	4, 22	.530	.13
<i>CFI $\times$ IF $\times$ CF</i>	0.39	4, 22	.811	.07
<i>Modality $\times$ Predictability $\times$ CFI $\times$ IF</i>	2.28	2, 24	.124	.16
<i>Modality $\times$ Predictability $\times$ CFI $\times$ CF</i>	2.21	2, 24	.131	.16
<i>Modality $\times$ Predictability $\times$ IF $\times$ CF</i>	0.27	4, 22	.892	.05
<i>Predictability $\times$ CFI $\times$ IF $\times$ CF</i>	0.36	4, 22	.835	.06
<i>Modality $\times$ Predictability $\times$ CFI $\times$ IF $\times$ CF</i>	0.32	4, 22	.861	.06

Notes. *Df* = Degrees of freedom. Bold italic values indicate two-tailed, italic values one-tailed significance.

Table 10 Mean RT (in ms; error rates in %, in parentheses) as a function of target modality, predictability, crossmodal feature identity, intramodal feature congruency (IF) and crossmodal feature compatibility (CF) in Experiment 3

	Vision				Audition			
Predictable blocks	IF				IF			
CF								
Crossmodal audiovisual	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	<i>Mean CF</i>	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	<i>Mean CF</i>
<i>Compatible</i>	683 (4.8)	696 (7.2)	705 (6.7)	695 (6.3)	707 (4.3)	787 (10.6)	833 (19.2)	776 (11.4)
<i>Neutral</i>	689 (2.9)	705 (6.7)	704 (8.7)	699 (6.1)	711 (2.9)	799 (10.1)	841 (24.0)	784 (12.3)
<i>Incompatible</i>	680 (4.8)	701 (8.2)	707 (5.8)	696 (6.3)	736 (5.3)	786 (18.8)	826 (20.7)	783 (14.9)
<i>Mean IF</i>	684 (4.2)	701 (7.4)	705 (7.1)		718 (4.2)	791 (13.1)	833 (21.3)	
Crossmodal tactile								
<i>Compatible</i>	676 (6.3)	697 (6.7)	717 (6.7)	697 (6.6)	759 (6.2)	801 (13.5)	824 (22.6)	795 (14.1)
<i>Neutral</i>	662 (5.8)	688 (6.7)	706 (6.7)	686 (6.4)	745 (6.7)	783 (12.0)	819 (17.8)	782 (12.2)
<i>Incompatible</i>	668 (3.8)	709 (7.2)	700 (8.2)	692 (6.4)	758 (5.8)	775 (13.9)	821 (16.3)	784 (12.0)
<i>Mean</i>	669 (5.3)	698 (6.9)	708 (7.2)		754 (6.3)	786 (13.1)	821 (18.9)	
Unpredictable blocks	IF				IF			
Crossmodal audiovisual	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	<i>Mean CF</i>	<i>Congruent</i>	<i>Neutral</i>	<i>Incongruent</i>	<i>Mean CF</i>
<i>Compatible</i>	754 (7.7)	771 (9.1)	781 (7.7)	779 (8.2)	846 (10.1)	884 (17.8)	890 (27.4)	873 (18.4)
<i>Neutral</i>	781 (14.4)	788 (11.1)	794 (10.6)	788 (12.0)	831 (5.3)	874 (19.7)	895 (20.7)	867 (15.2)
<i>Incompatible</i>	771 (17.8)	793 (12.0)	816 (11.5)	794 (13.8)	880 (9.6)	873 (21.2)	919 (19.7)	890 (16.8)
<i>Mean</i>	779 (13.3)	784 (10.7)	798 (9.9)		852 (8.3)	877 (19.6)	901 (22.6)	
Crossmodal tactile								
<i>Compatible</i>	779 (8.2)	786 (5.8)	784 (11.1)	783 (8.3)	846 (11.6)	879 (25.5)	869 (21.2)	865 (19.4)
<i>Neutral</i>	758 (11.5)	773 (8.2)	777 (8.2)	769 (9.3)	870 (7.7)	878 (18.8)	874 (21.6)	874 (16.0)
<i>Incompatible</i>	778 (11.5)	789 (8.7)	789 (9.1)	786 (9.8)	867 (6.7)	862 (17.8)	890 (22.6)	873 (15.7)
<i>Mean IF</i>	772 (10.4)	783 (7.5)	784 (9.5)		861 (8.7)	873 (20.7)	878 (21.8)	

References

- Chan, J. S., Merrifield, K., & Spence, C. (2005). Auditory spatial attention assessed in a flanker interference task. *Acta Acustica united with Acustica*, 91, 554–563.
- Driver, J., & Grossenbacher, P. G. (1996). Multisensory spatial constraints on tactile selective attention. In I. Toshio & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 209–235). Cambridge, MA: MIT Press.
- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1–11.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143–149.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1030–1044.
- Fong, M. C. M., Hui, N. Y., Fung, E. S. W., Chu, P. C. K., & Wang, W. S. Y. (2018). Conflict monitoring in multi-sensory flanker tasks: Effects of cross-modal distractors on the N2 component. *Neuroscience Letters*, 670, 31–35.
- Gallace, A., Soto-Faraco, S., Dalton, P., Kreukniet, B., & Spence, C. (2008). Response requirements modulate tactile spatial congruency effects. *Experimental Brain Research*, 191, 171–186.
- Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 699–706.
- Hommel, B., Memelink, J., Zmigrod, S., & Colzato, L. S. (2014). Attentional control of the creation and retrieval of stimulus-response bindings. *Psychological Research*, 78, 520–538.
- Huang, W., Su, Y., Zhen, Y., & Qu, Z. (2016). The role of top-down spatial attention in contingent attentional capture. *Psychophysiology*, 53, 650–662.
- Ito, M., & Kawahara, J. I. (2016). Contingent attentional capture across multiple feature dimensions in a temporal search task. *Acta Psychologica*, 163, 107–113.
- Jensen, A., Merz, S., Spence, C., & Frings, C. (2019). Overt spatial attention modulates multisensory selection. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 174–188.
- Klapetek, A., Ngo, M. K., & Spence, C. (2012). Do crossmodal correspondences enhance the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, 74, 1154–1167.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human, Perception and Performance*, 21, 451–468.

- Lavie, N. (1997). Visual feature integration and focused attention: Response competition from multiple distractor features. *Perception & Psychophysics*, 59, 543–556.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9, 75–82.
- Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, 19, 143–148.
- Lunn, J., Sjöblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement of attention depends on whether you are already paying attention. *Cognition*, 187, 38–49.
- Mast, F., & Frings, C. (2014). The impact of the irrelevant: The task environment modulates the impact of irrelevant features in response selection. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 2198–2213.
- Mast, F., Frings, C., & Spence, C. (2014). Response interference in touch, vision, and crossmodally: Beyond the spatial dimension. *Experimental Brain Research*, 232, 2325–2336.
- Mast, F., Frings, C., & Spence, C. (2015). Multisensory top-down sets: Evidence for contingent crossmodal capture. *Attention, Perception, & Psychophysics*, 77, 1970–1985.
- Mast, F., Frings, C., & Spence, C. (2017). Crossmodal attentional control sets between vision and audition. *Acta Psychologica*, 178, 41–47.
- Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, 18, 904–909.
- Matusz, P. J., & Eimer, M. (2013). Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*, 50, 996–1009.
- Maxwell, S. E., & Delaney, H. D. (2004). Designing experiments and analyzing data: A model comparison perspective (2nd ed.). New York, NY: Psychology Press.
- Merz, S., Jensen, A., Spence, C., & Frings, C. (2019). Multisensory distractor processing is modulated by spatial attention. *Journal of Experimental Psychology: Human Perception & Performance*. Advance online publication. doi:<https://doi.org/10.1037/xhp0000678>
- Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech. *Information Fusion*, 11, 4–11.
- Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science*, 12, 511–515.
- Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, 15, 398–403.
- Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1311–1321.
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, 1191, 182–200.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73, 971–995.
- Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: Reflexive crossmodal orienting and ventriloquism. *NeuroReport*, 11, 2057–2061.
- Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, 63, 330–336.
- Spence, C., Ranson, J., & Driver, J. (2000). Crossmodal selective attention: Ignoring auditory stimuli presented at the focus of visual attention. *Perception & Psychophysics*, 62, 410–424.
- Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research*, 258, 134–142.
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, 14, 400–410.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 799–806.
- Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison Wesley.
- Wada, Y. (2011). Crossmodal contextual effects of tactile flankers on the detection of visual targets. *i-Perception*, 2, 846.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*, 108, 21–33.
- Wang, B., & Theeuwes, J. (2018). Statistical regularities modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 44, 13–17.
- Wesslein, A. K., Spence, C., & Frings, C. (2014). When vision influences the invisible distractor: Tactile response compatibility effects require vision. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 763–774.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.